Euphytica **128:** 261–267, 2002.

Analysis of a single nucleotide polymorphism that controls the cooking quality of rice using a non-gel based assay

Concetta A. Bormans[1], Richard B. Rhodes[2,4], Daniel D. Kephart[2], Anna M. McClung[3] & William D. Park[1,*]

[1]*Borlaug Center for Southern Crop Improvement, Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843, U.S.A.;* [2]*Promega Corporation, 2800 Woods Hollow Road, Madison, WI 53711, U.S.A.;* [3]*USDA-ARS, 1509 Aggie Dr., Beaumont, TX 77713, U.S.A.;* [4]*Present address: Minnesota Molecular and Cellular Therapeutics, University of Minnesota, Minneapolis, MN 55455, U.S.A.; (*author for correspondence, e-mail: wdpark@tamu.edu)*

Received 22 July 2001; accepted 15 April 2002

Key words: amylose, granule-bound starch synthase, READIT™ assay, single-nucleotide polymorphism, waxy gene

Summary

The waxy gene encoding granule-bound starch synthase (GBSS) is responsible for the synthesis of amylose in developing grain. Recent work has shown that a G-T polymorphism in the leader intron 5' splice site of GBSS plays a key role in determining the cooking and processing quality of rice. Cultivars with sequence AGGTATA at this location splice GBSS pre-mRNA efficiently and produce relatively large amounts of amylose. These varieties generally a have firm texture when cooked and the grains remain separate. In contrast, GBSS pre-mRNA splicing is temperature sensitive and generally less efficient in cultivars with the sequence AGTTATA. As a result, these cultivars generally have lower amylose content and produce soft and sticky cooked rice. We have used the READIT™ assay, a non-gel based assay that uses the ability of DNA polymerase to perform pyrophosphoralysis, the reverse of DNA polymerization, to screen the critical G-T polymorphism in more than 750 samples from U.S. and Asian germplasm. We observed complete concordance between the results obtained using DNA sequencing or restriction enzyme digestion and the READIT™ assay. It also gave accurate results with both heterozygous plants and with complex mixtures as might result when grain from advanced generation plants is pooled to obtain larger samples.

Abbreviations: GBSS – granule-bound starch synthase; PMP – paramagnetic particle; SNP – single-nucleotide polymorphism; RLU – relative light unit; RRR – relative response ratio

Introduction

Amylose content is the single most important characteristic for predicting the cooking and processing quality of rice (Juliano, 1985). Varieties with low amylose are typically soft and sticky, while varieties with intermediate and high amylose tend to be firm and dry, and remain separate after cooking.

The waxy gene encoding granule bound starch synthase (GBSS) is primarily responsible for amylose synthesis in rice as in other cereals (Sano, 1984). At least eight different alleles of GBSS in rice can be distinguished based on a CT repeat and a single nucleotide polymorphism (SNP) at the 5' leader intron splice site. These alleles accounted for more than 85% of the variation in amylose content in an extended pedigree of 89 US rice cultivars (Ayres et al., 1997). Strikingly, all of the varieties examined in this study which had amylose contents of more than 18% were found to have the sequence AGGTATA at the 5' splice site, while those with less than 18% amylose had the sequence AGTTATA as this position.

Subsequent work has shown that this G-T polymorphism plays a key role in GBSS pre-mRNA splicing. Varieties with the sequence AGGTATA at this location splice efficiently and accumulate large amounts of only the mature GBSS mRNA. In contrast, varieties with the sequence AGTTATA generally have lower levels of mature GBSS mRNA, accumulate GBSS pre-mRNA that still contains the leader intron, and utilize a variety of alternate splice sites (Bligh et al., 1998). Interestingly the effect of the G-T polymorphism is temperature dependent. At 18 °C, this polymorphism has little effect on the steady state level of GBSS mRNA, while at 25 °C or 32 °C, the presence of a 'T' at the critical position dramatically reduces the accumulation of mature GBSS mRNA. The variety 'Toro-2', for example, has 20.8% amylose when grown at 18 °C. When the temperature was increased to 32 °C, the amylose content decreased to 11.9% (Larkin & Park, 1999).

The temperature sensitivity of amylose synthesis represents a problem for traditional methods of assaying rice quality. If a particular breeding plot happens to mature under unusual environmental conditions, its amylose content may not reflect the breeding lines' typical grain quality. Rather than trying to predict amylose content across years and locations based on a single field plot, it would be much better to directly assay the specific DNA differences that determine amylose content and thus grain quality. Directly assaying the G-T polymorphism would also avoid the problem of dominance and allow heterozygous plants to be accurately distinguished from homozygotes in early generation screening.

For use in marker-assisted selection in a breeding program, mutation detection assays must meet several criteria. Allelic scoring and must be reproducible and reliable, and the ability to detect heterozygotes is essential. Perhaps most important, assays must be fast, inexpensive, and at least partially amenable to automation to allow for high throughput screening. A plant breeder can have thousands of progeny each year, which must be screened for multiple traits. All of this must be done within a budget that requires a much lower price per sample than what is customary in clinical diagnostic settings. Because of this, any assay employed in a breeding program must be easy to develop and test while requiring the minimum amount of cost associated with new probe design and testing. The use of standard oligonucleotides and common reagent formats for a detection of a wide variety of sequence variations is an additional benefit.

The READIT™ assay is a novel, non-gel based assay for DNA mutation detection (Promega Corporation, Madison, WI). This assay takes advantage of the fact that in the presence of high concentrations of pyrophosphate, the DNA polymerase reaction can perform the reverse of DNA polymerization to produce dNTPs. Coupled enzymes use the dNTPs that are generated to convert ADP to ATP, which is then quantified using luciferase. In the absence of 3' to 5' exonuclease activity, the ability of DNA polymerase to move backwards is completely dependent upon complementary pairing between interrogation oligonucleotides specific for individual alleles and sample DNA (Figure 1).

Materials and methods

PCR analysis

To compare the READIT™ Technology assay to conventional assays, we first analyzed 182 U.S. and Asian rice cultivars representing long, short, and medium grains, as well as both the *indica* and *japonica* subspecies. Within the *japonica* subspecies, we included both temperate *japonica* and *javonica* cultivars. Samples included high, medium and low amylose cultivars, as well as one glutinous cultivar that contains essentially no amylose. In addition, we also utilized the assay in a more diagnostic fashion by screening 576 varieties obtained from the USDA-ARS National Small Grains Collection previously reported to have low amylose content. For sample analysis, approximately 20 ng of purified genomic DNA was amplified by Polymerase Chain Reaction (PCR) using 50 pmol of upstream primer (5'-TCTCAAGACACAAATAACTGCAG) and 50 pmol downstream primer (5'-CCCAACACCTTACAGAAATTAGC) designed to amplify a 240 bp (bases 25–265, Genbank Accession AF031162) segment of the rice granule-bound starch synthase (*Waxy*) gene. The downstream primer was synthesized to contain 3 consecutive phosphorothioate linkages at the 5' end. PCR reactions contained 1.25 units Taq DNA Polymerase (Promega), 1.5 mM $MgCl_2$, and 200 μM dNTP mix. Thermocycling was performed under the following conditions; 1 cycle of 2 min at 94 °C; 35 cycles of 0.5 min at 94 °C, 1 min at 60 °C, and 1 min at 70 °C; 1 cycle of 5 min at 70 °C.

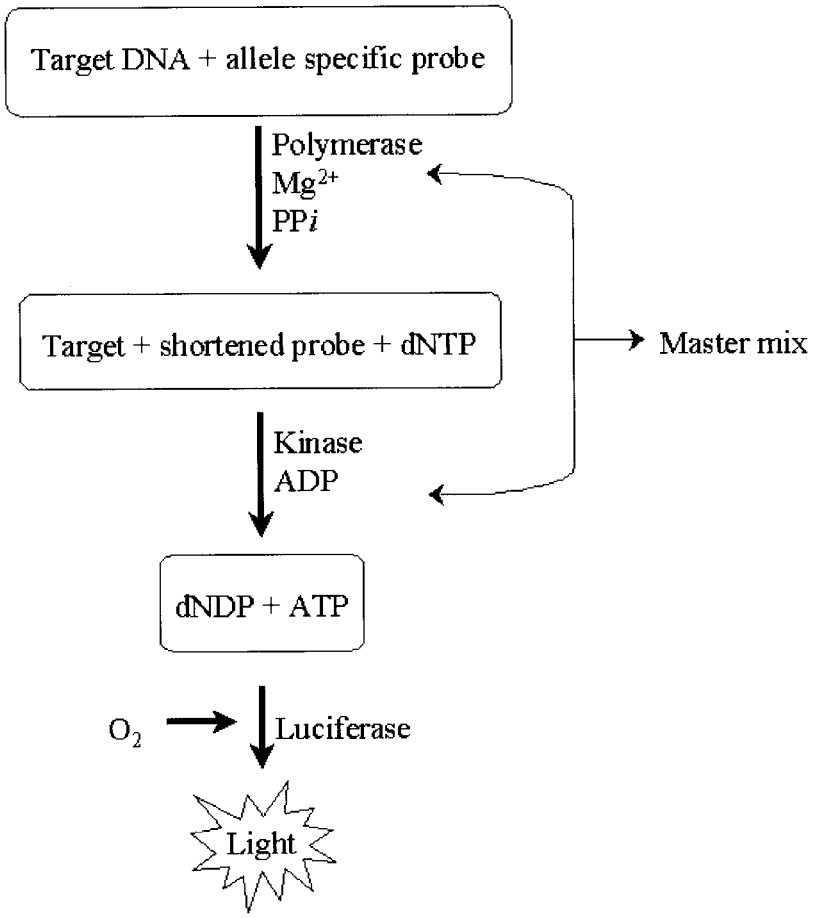

Figure 1. Flow chart demonstrating the READIT™ assay. The target DNA is denatured and master mix containing the following is added: allele specific probe, polymerase, magnesium, pyrophosphate, ADP, and kinase. The reaction results in the release of dNTPs from the probe and subsequent formation of ATP by the kinase. Addition of luciferase results in the conversion of ATP to light, which is then detected using a luminometer.

PCR product purification

Aliquots of each PCR reaction (25 μl) were treated with 12.5 units of T7 Gene 6 exonuclease (United States Biochemical, Cleveland, OH) for 30 minutes at 37 °C. Nuclease treated amplification products were then purified using MagneSil™ paramagnetic particles (PMPs) (Promega) on a Biomek® 2000 robotics workstation (Beckman Coulter, Inc., Fullerton, CA) using robotics protocols provided by the manufacturer. Briefly, 150 μl of the PMPs in guanidine thiocyanate binding buffer were added to each sample in the wells of a 96-well plate. Samples were incubated for 2 minutes at room temperature, and the PMPs were rapidly captured using a 96-well magnetic pin array (Promega). PMPs in each well were washed with three 150 μl exchanges of 70% ethanol. Water was added to each well (50 μl), the PMPs were resuspended, and 150 μl of binding buffer were added to facilitate a second capture of the PCR products onto the same bed of PMPs. After magnetic capture of the PMPs, three additional 150 μl washes of 70% ethanol were performed. Purified samples were eluted in 100 μl of nuclease-free water.

READIT™ Technology Interrogation

Automated READIT™ assay reactions were performed using a Biomek 2000 instrument. Briefly, 5 μl of the purified PCR sample were combined with

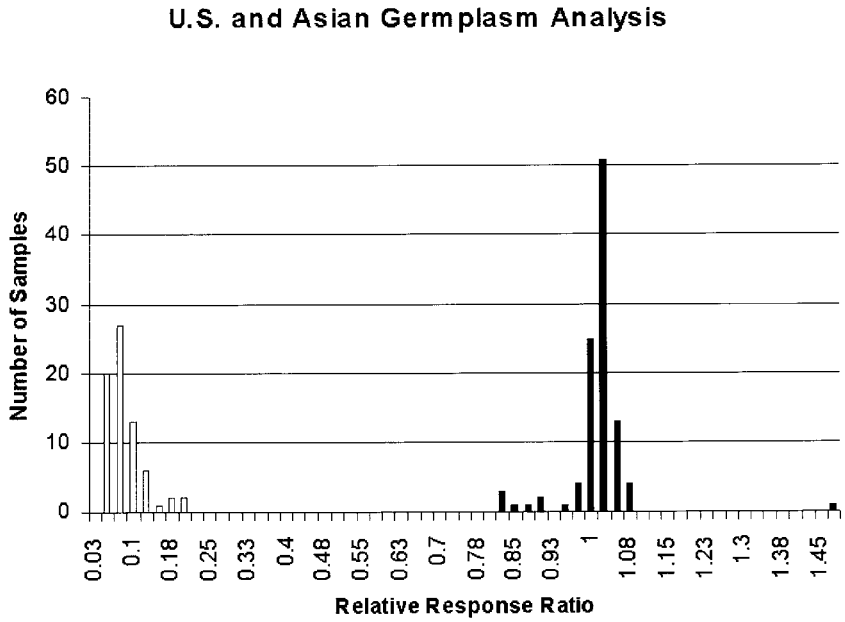

Figure 2. READIT™ Technology analysis of U.S. and Asian germplasm. 182 genomic DNA samples representing a wide variety of U.S. and Asian germplasm were analyzed with the READIT™ assay. Using probes directed toward the GBSS gene SNP, each sample was characterized for the presence of a 'G' or a 'T' at the SNP site. All samples were correctly called with the READIT™ assay when the data were compared to data previously obtained from amylose analysis. White bars indicate 'T' alleles and black bars indicate 'G' alleles.

5 μl of 0.06 N NaOH and samples were incubated at room temperature for 5 minutes. Ten μl of neutralization solution (50 mM Tris (pH 7.3), 10 mM $MgCl_2$) that contained 150 pmol of 'G' specific probe (5' – AGAACATCTGCAAGG), or 'T' specific probe (5' – AGAACATCTGCAAGT) were added. Neutralization solution alone was added to a third tube to serve as reaction background. Chemically denatured samples were incubated for 5 minutes at 37° C and 25 μl of READIT™ version 1.0 System Master Mix were injected into each sample. The READIT™ version 1.0 Master Mix contained 1 × DNA polymerase buffer (Promega), 0.1 U/μl Klenow exonuclease minus (Promega), 2 mM sodium pyrophosphate, 0.2 μM ADP, and 0.01 U/μl nucleoside diphosphate kinase (Sigma, St. Louis, MO). After addition of Master Mix, samples were incubated for 15 minutes at 37 °C. One hundred μl of L/L Reagent (Promega) were injected per well and Relative Light Units (RLUs) for each well were integrated for 10 seconds in a EG&G LB 96V luminometer (Perkin Elmer, Gaithersberg, MD).

The READIT™ assay was compared to a conventional gel based method for the evaluation of heterozygotes. Genomic DNA samples of the varieties 'Lemont' and 'Teqing', which have the 'G' and 'T' alleles, respectively, were mixed in the following ratios: 30/70% G/T, 50/50% G/T, 60/40% G/T, 80/20% G/T, and 90/10% G/T. Reactions were amplified in parallel reactions as described above. In one set of amplifications, 10 μCi of ^{32}P dCTP were added. The resulting radiolabeled PCR products were digested with the restriction enzyme *Acc*I. One microliter of 25 mM $MgCl_2$ and 10 units of *Acc*I were added to 10 μl of PCR product. The digestions proceeded at 37 °C for 1 h. The digested products were run on a 10% acrylamide gel. The gel was dried for 1 h at 65 °C, followed by exposure to a Fluorimager® screen (Molecular Dynamics, Sunnyvale, CA) for 1 h. For each heterozygote sample, the exposed areas on the screen representing the digested (120 bp) and undigested (240 bp) PCR product were integrated. Ratios of the digested and undigested counts to total counts were then calculated.

Data analysis was performed using the READIT™ Technology Calculator© software provided by Promega Corporation. Relative light units (RLUs) obtained from 'G' and 'T' interrogation reactions were adjusted by subtracting background RLUs obtained in the absence of interrogation probe. A Relative Response Ratio (RRR) was calculated by dividing the

adjusted 'G' allele RLUs by the summed RLUs from both 'G' and 'T' alleles. Since there was a greater than 6 standard deviation separation between the RRRs of each allele, in some cases genotypes were called automatically by setting the calling windows as three standard deviations from the mean of each genotype population.

Results

The G-T polymorphism in the GBSS gene of 182 rice varieties, representing a wide range of U.S. and Asian germplasm, were analyzed by direct sequencing or restriction enzyme digest with *Acc*I. Results from the READIT^{TM} assay agreed 100% with sequencing and *Acc*I digest. As shown in Figure 2, there was a large difference in the relative response ratio of homozygous 'G' and 'T' alleles in both the U.S. and Asian samples, which allowed unambiguous allele assignments.

To determine the assay's sensitivity in detecting heterozygotes, varieties homozygous for the 'G' or 'T' allele were also mixed in known ratios. A linear relationship (R^2 = 0.99) was observed between the RRR obtained and the allelic frequency observed in samples of artificial mixtures as determined by restriction enzyme digestion. For heterozygote calling of single samples, a calling window is defined and samples that fall within the defined range are called as heterozygotes. However, READIT^{TM} analysis of samples with known ratios of the 'T' and 'G' alleles accurately determined the frequencies of these two alleles (Figure 3). This indicates that the assay is linear under our conditions.

The extreme accuracy of the assay allowed us to use it to screen for mutants having the 'G' allele and inefficient splicing, resulting in low amylose. Low amylose varieties from the National Small Grains Collection were assayed as described. However, since varieties with low amylose are expected to have only the 'T' allele, those with a relative response ratio of 0.3 or greater were flagged for further analysis (Figure 4). Twenty of the NSGC low amylose varieties were called homozygous for the 'G' allele and were confirmed by DNA sequencing. RVA analysis of each homozygous 'G' variety revealed that all but three were high amylose as expected from this genotype. The three low amylose 'G' varieties will be subjected to further analysis.

Discussion

Single nucleotide polymorphisms, though used extensively as markers in animal and human research, have not yet been developed to such a degree in agricultural research. However, current genome sequencing projects for *Arabidopsis*, rice, maize, wheat, and other crop species are providing a wealth of new sequencing data, which will serve to accelerate SNP discovery. SNPs linked to disease resistance genes have been identified in soybean and common bean (*Phaseolus vulgaris* L.) (Meksem et al., 2001; Melotto & Kelly, 2001), and it has been shown that the DNA of chromosome 1 in maize has a higher ratio of SNPs per base pairs than that of both *Drosophila melanogaster* and humans (Tenaillon et al., 2001). Clearly, a robust and reliable system for SNP detection is needed. Current methods using gel electrophoresis are time consuming and labor intensive unless one has access to expensive automated electrophoresis apparatus. Newer methods such as mass spectrometry and molecular beacons can be automated, but they also require expensive, specialized equipment and reagents (Haff & Smirnov, 1997; Tyagi et al., 1997). The READIT^{TM} assay is a high throughput, non-gel based system that can be performed using relatively inexpensive luminometers, standard enzymes and oligonucleotide primers available from several companies.

Our results demonstrate a very clear separation of GBSS alleles in a very wide range of germplasm (Figure 2). We also found that the assay can accurately detect not only 50/50% heterozygotes, but also the wide range of allelic frequencies found in artificial mixtures or F_5 derived, F_3 pools (data not shown). The ability to accurately score complex heterozygotes is extremely useful in breeding for rice grain quality, since progeny from many plants must often be pooled to obtain sufficient material for processing and sensory analysis. The ability to accurately score other than 50/50% heterozygotes would also be useful for SNP detection in polyploid genomes. Additionally, detection of more than two alleles at a single locus can be done by the addition of another interrogation primer to the assay.

Another important feature for a high throughput genotype assay is the ability to distinguish a failed reaction and not misscore it. For example, in scoring the G-T polymorphism using *Acc*I, failure of the *Acc*I digest could falsely indicate the presence of the 'T' allele. Since scores are determined by dividing the Relative Light Units (RLUs) of the corrected 'G' allele

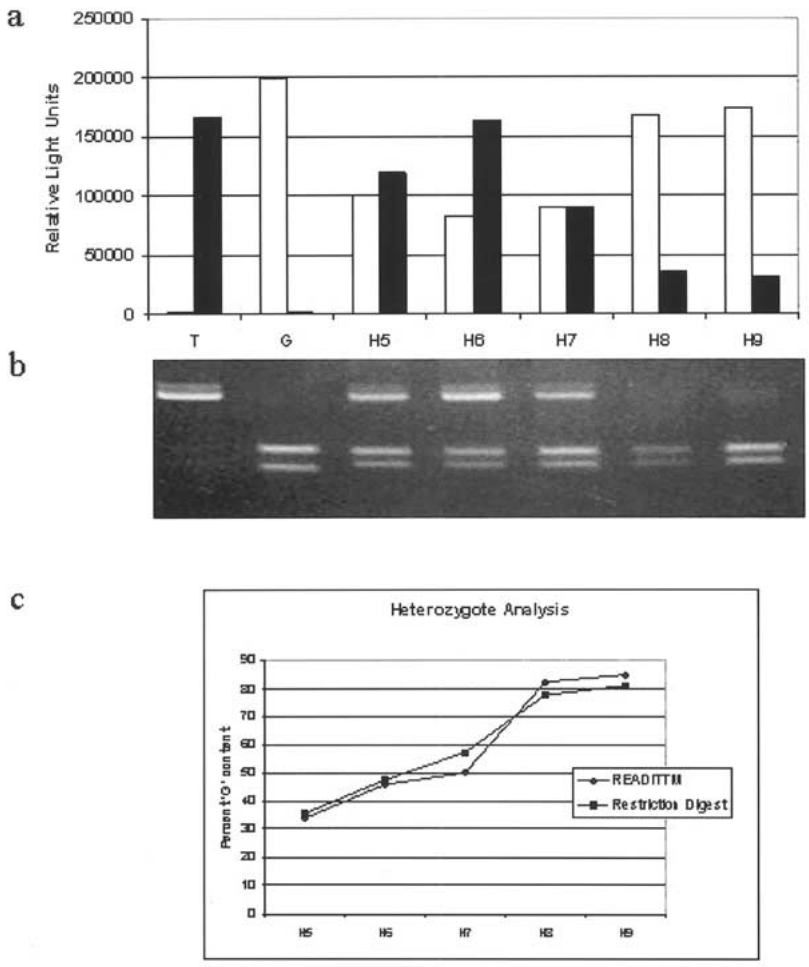

Figure 3. Comparison of the READIT^{TM} assay with conventional restriction enzyme digest assays. (a) Templates containing the 'G' or 'T' base at the SNP position were prepared as described in the text and analyzed with both the READIT^{TM} assay and *Acc*I restriction enzyme digests. The mixtures of templates were probed with both the 'G' and 'T' interrogation oligonucleotides and the percentage of 'G' (white bars) or 'T' template (black bars) in each mixture was then determined with the READIT^{TM} assay. (b) The same samples were analyzed in a PCR reaction containing ^{32}P-dCTP followed by PAGE and scanning densitometry. (c) Accuracy of the READIT^{TM} assay and conventional restriction digest assays for identification of heterozygotes. The percent 'G' allele (Y-axis) as determined by both methods was compared for each of the heterozygote samples H5-H9.

by the total RLUs from both alleles to obtain a relative response ratio, failed reactions are automatically marked as candidates for reevaluation if the RLUs are not significantly higher than the background signal. In a typical reaction of a 'G' homozygote, we obtained 458,265 RLUs with the 'G' specific oligonucleotide, 95,890 RLUs with the 'T' specific oligonucleotide, and 92,591 RLUs for background. For a 'T' homozygote, we typically obtained 141,633 RLUs with the 'G' specific oligonucleotide, 575,781 RLUs with the 'T' specific oligonucleotide, and 137,968 RLUs for background. Theoretically, the optimum response ratios for each allele are 0 and 1. However, when the RLUs of one allele are less than that of the background reaction, it is possible to obtain ratios outside these limits, as can be seen with the low amylose varieties obtained from the NSGC (Figure 4). The ability to set limits for each allele, as well as for heterozygotes, allows rapid genotyping while singling out individuals that may require further analysis. For example, of the 576 low amylose varieties, only 20 had relative response ratios of greater than 0.75, classifying them as homozygous for the 'G' allele. The genotypes of those samples were then confirmed by DNA

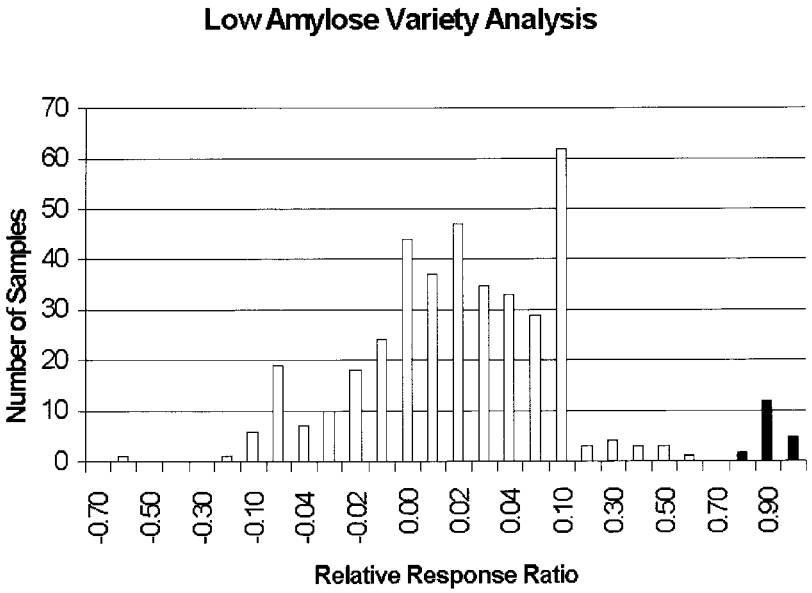

Figure 4. READIT^{TM} Technology analysis of low amylose varieties. 576 varieties that have been classified as low (<18%) amylose were obtained from the National Small Grains Collection and analyzed with the READIT^{TM} assay. As expected, most samples were scored as homozygous for the 'T' allele (white bars). Samples having a Relative Response Ratio (RRR) between 0.33 and 0.66 were called "ambiguous' (gray bars). Samples with a RRR greater than 0.75 were called homozygous for the 'G' allele (black bars). All of the ambiguous and homozygous 'G' varieties were confirmed by DNA sequencing, and subsequent amylose assays showed that all but three of them in fact had high amylose contents as expected with this genotype.

sequencing. The amylose contents of those individuals were assayed and all but three were high amylose varieties. This again demonstrated the advantage of measuring genotype rather than phenotype. It is possible that previous measurements were affected by growth conditions such as temperature.

In summary, this technique provides a method for detecting a variety of DNA polymorphisms without requiring large amounts of sample, costly instrumentation, reagents, or large amounts of time for analysis.

Acknowledgements

We thank Grace Walker for her invaluable assistance in sample preparation and organization. This research was supported by a grant from the Texas Grain and Grass Initiative.

References

Ayres, N.M., A.M. McClung, P.D. Larkin, H.F.D. Bligh, C.A. Jones & W.D. Park, 1997. Microsatellites and a single nucleotide polymorphism differentiate apparent amylose classes in an extended pedigree of US rice germplasm. Theor Appl Genet 94: 773–781.

Bligh, H.F.J., P.D. Larkin, P.S. Roach, C.A. Jones, H. Fu & W.D. Park, 1998. Use of alternate splice sites in granule-bound starch synthase mRNA from low-amylose varieties. Plant Molec Biol 38: 407–415.

Haff, L. & I. Smirnov, 1997. Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. Genome Meth 7: 378–388.

Juliano, B.O., 1985. Rice Chemistry and Technology. American Association of Cereal Chemists, St Paul, MN.

Larkin, P.D. & W.D. Park, 1999. Transcript accumulation and utilization of alternate and non-consensus splice sites in rice granule-bound starch synthase are temperature-sensitive and controlled by a single-nucleotide polymorphism. Plant Molec Biol 40: 719–727.

Meksem, K., E. Ruben, D.L. Hyten, M.E. Schmidt & D.A. Lightfoot, 2001. High-throughput genotyping for a polymorphism linked to soybean cyst nematode resistance gene Rhg4 by using Taqman^{TM} probes. Molec Breed 7: 63–71.

Melotto, M. & J.D. Kelly, 2001. Fine mapping of the Co-4 locus of common bean reveals a resistance gene candidate, COK-4, that encodes a protein kinase. Theor Appl Genet 103: 508–517.

Sano, Y., 1984. Differential regulation of waxy gene expression in rice endosperm. Theor Appl Genet 68: 467–473.

Tenaillon, M.I., M.C. Sawkins, A.D. Long, R.L. Gaut, J.F. Doebley & B.S. Gaut, 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp mays L.). Proc Nat Acad Sci 98: 9161–9166.

Tyagi, S., D. Bratu & F. Kramer, 1998. Multicolor molecular beacons for allele discrimination. Nature Biotech 16: 49–53.